Molecular mechanics calculations on Rous sarcoma virus protease with peptide substrates

IRENE T. WEBER AND ROBERT W. HARRISON

Department of Microbiology and Immunology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

(RECEIVED February 10, 1997; ACCEPTED June 30, 1997)

Abstract

Molecular models of Rous sarcoma virus (RSV) protease and 20 peptide substrates with single amino acid substitutions at positions from P4 to P3', where the scissile bond is between P1 and P1', were built and compared with kinetic measurements. The unsubstituted peptide substrate, Pro-Ala-Val-Ser-Leu-Ala-Met-Thr, represents the NC-PR cleavage site of RSV protease. Models were built of two intermediates in the catalytic reaction, RSV protease with peptide substrate and with the tetrahedral intermediate. The energy minimization used an algorithm that increased the speed and eliminated a cutoff for nonbonded interactions. The calculated protease–substrate interaction energies showed correlation with the relative catalytic efficiency of peptide hydrolysis. The calculated interaction energies for the 8 RSV protease–substrate models with changes in P1 to P1' next to the scissile bond gave the highest correlation coefficient of 0.79 with the kinetic measurements, whereas all 20 substrates showed the lower, but still significant correlation of 0.46. Models of the tetrahedral reaction intermediates gave a correlation of 0.72 for the 8 substrates with changes next to the scissile bond, whereas a correlation coefficient of only 0.34 was observed for all 20 substrates. The differences between the energies calculated for the tetrahedral intermediate and the bound peptide gave the most significant correlation coefficients of 0.90 for models with changes in P1 and P1', and 0.56 for all substrates. These results are compared to those from similar calculations on HIV-1 protease and discussed in relation to the rate-limiting steps in the catalytic mechanism and the entropic contributions.

Keywords: aspartic proteases; calculated interaction energies; RSV protease; substrate binding; transition state

The Rous sarcoma virus (RSV) protease is a member of the aspartic protease family and is enzymatically active as a dimer. The crystal structure of RSV protease (Miller et al., 1989; Jaskolski et al., 1991) has a three-dimensional fold similar to that of HIV-1 protease (Wlodawer et al., 1989; Weber, 1990). Although there are many crystal structures of HIV protease with different inhibitors (for review, see Wlodawer & Erickson, 1993), there is no crystal structure of RSV protease in the complex with an inhibitor. Instead, the structure of the RSV protease–substrate complex has been modeled by analogy to the structures of HIV protease with peptide-like inhibitors, as described in Grinde et al. (1992). RSV and HIV-1 proteases hydrolyze different cleavage sites in their viral polypeptide substrates (Oroszlan & Luftig, 1990), and RSV protease does not cleave peptides representing the natural substrates of HIV-1 protease (Cameron et al., 1992). The molecular basis of substrate specificity of RSV and HIV-1 proteases is not obvious, due to the lack of a clear consensus sequence in the different cleavage sites. The substrate specificity of RSV protease

has been analyzed by kinetic measurements of hydrolysis of peptides that represent the natural cleavage sites, and variations of these peptides with different amino acids substituted at each position from P3 to P3' in the standard nomenclature, where the peptide bond between P1 and P1' is cleaved (Cameron et al., 1993; Tozser et al., 1996). Cameron et al. (1993) tested substitutions of the peptide, Pro-Pro-Ala-Val-Ser-Leu-Ala-Met-Thr-Met-Arg-Arg, that represented the NC-PR cleavage site of RSV protease. This peptide is a good substrate for both RSV and HIV-1 proteases. It is unusual in having relatively small amino acid side chains, Ala, Ser, and Val, at P3 to P1 and P2'. The prevalence of smaller side chains in this peptide means that fewer conformations are accessible for the side chains and entropic effects may be reduced, which is an advantage for the energy calculations. The peptides tested by Cameron et al. (1993) were chosen to have substitutions at each position from P4 to P3' of the substrate peptide that included different sizes and types of amino acid side chains. Twenty of these peptide substrates were modeled in the complex with RSV protease.

Recently, molecular mechanics calculations have proven valuable for estimating the free energy difference between two states related by small structural changes. The molecular mechanics interaction energy for HIV protease with different inhibitors showed

Reprint requests to: I.T. Weber, BLSB Rm. 833, Department of Microbiology and Immunology, Kimmel Cancer Center, Thomas Jefferson University, 233 South 10th St., Philadelphia, Pennsylvania 19107; e-mail: weber@asterix.jci.tju.edu.

high correlation with the measured inhibition in calculations of Holloway et al. (1995) using the MM2X force field. Our calculations with the program AMMP (Harrison, 1993) have shown significant correlation with free energy differences derived from kinetic data for novel inhibitors of trypsin (Kourinov & Harrison, 1994), modeling the structure and energetics of glucokinase with different sugar substrates (Xu et al., 1995), and peptide substrates of HIV-1 protease (Weber & Harrison, 1996). The calculated protein-ligand interaction energies were shown to correlate with the free energy differences derived from kinetic measurements, despite the absence of bulk water in the calculation. These relatively short but accurate molecular mechanics calculations save between 10- and 100-fold cpu time per point compared to the longer simulations required for free energy perturbation estimations of the differences in free energy. The fundamental improvement in accuracy of AMMP (Harrison, 1993) is due to the inclusion of all long-range non-bonded energies and all hydrogen atoms. Tests of these improvements are described in Harrison and Weber (1994), Kourinov and Harrison (1994), and Harrison et al. (1995).

In order to understand the energetics of peptide hydrolysis by RSV protease, we have run molecular mechanics calculations on several different substrates. Similar calculations were run previously on HIV protease with peptide substrates (Weber & Harrison, 1996). The calculations were run on models that represent two intermediates in the AAC2 reaction mechanism, which is the predominant form seen in acid-catalyzed hydrolysis of amides in model systems (March, 1992). The two reaction intermediates that were modeled were RSV protease with bound peptide substrate and with the tetrahedral reaction intermediate. Models were built of RSV protease with a series of 20 related peptide substrates, and the protease-substrate interaction energies were calculated in AMMP. The energies of solvation from bulk solvent and the entropy terms were not calculated. The interaction energies of the free peptide substrates with water were assumed to be constant. The RSV protease-substrate interactions in the models are compared to those observed in HIV-1 protease-inhibitor crystal structures (Wlodawer & Erickson, 1993; Gustchina et al., 1994). The calculated interaction energies are compared to differences in free energy derived from kinetic measurements of Cameron et al. (1993). The observed correlations are discussed in relation to the slow step in the reaction mechanism and the effects of entropy.

Results and discussion

Rationale for the models of reaction intermediates

The retroviral proteases are thought to catalyze peptide hydrolysis by a general acid-general base mechanism without the formation of a covalent enzyme-substrate intermediate, as described for the nonviral aspartic proteases (Fruton, 1987). The catalytic mechanism of RSV protease has not been studied, although it is presumed to be similar to that of HIV protease. The mechanism of HIV protease has been investigated recently (Hyland et al., 1991; Polgar et al., 1994). The reaction mechanism included the binding of substrate to protease and the stabilization of a tetrahedral intermediate. Models were built of both these protease complexes with reaction intermediates (Fig. 1). The unliganded protease structure was not modeled.

The major problem in modeling a substrate in aspartic acid proteases is the placement of the hydrogens involved in the reaction. The optimal pH for catalysis is near the pK of aspartic acid,

RSV Protease Reaction Intermediates

Fig. 1. Scheme of the active site for the two reaction intermediates that have been modeled. The RSV protease model with the scissile peptide included a proton (H^+) between the two catalytic aspartates (Asp 37 and Asp 37') and the water molecule that is conserved in crystal structures of HIV protease with peptide-like inhibitors. The tetrahedral intermediate has $[C(OH)_2-NH_2]^+$ instead of the peptide group $CO-NH$ with an overall delocalized charge of +1.

so, on average, only one acid is protonated. Although there are experimental data showing that, in protease-inhibitor complexes, one aspartic acid is protonated (Smith et al., 1996), in studying the reaction kinetics, all of the possible configurations should be considered. The relevant configuration may not be the major configuration in the equilibrium context of protease with inhibitor. There are 5 single-protonated complexes for the two catalytic aspartic acids and the carbonyl oxygen of the scissile bond (Fig. 1) and 10 double-protonated complexes. If treated independently, these complexes would require up to 15 scaling parameters. Instead, we chose to use an average model that corresponds to a putative intermediate between discrete protonation states. The model contains a proton that interacts with both catalytic aspartic residues, Asp 37 and Asp 37', and the carbonyl oxygen of the scissile peptide bond. The proton is a molecular mechanics model of an average configuration for a system in exchange among five protonated states. This proton was stable during at least 40 ps of molecular dynamics simulation and it forms multicenter hydrogen bonds with all four Asp δ oxygen atoms and the carbonyl oxygen of the scissile bond. Because multicenter hydrogen bonds are common in inorganic complexes (Jeffrey & Saenger, 1991), this protonation model is physically valid.

The model for the tetrahedral intermediate consisted of $[C(OH)_2-NH_2]^+$ instead of the scissile peptide group $CO-NH$, the proton and the water molecule that interacted with the substrate or inhibitor, and the flaps (Fig. 1). A model with a protonated secondary amine $[C(OH)_2-NH_2]^+$ was chosen, because, with other protonated intermediates (such as $\{[C(OH)(H_2O)]^+-NH\}$), the highly basic secondary amine would extract the proton to form the protonated amine. The water molecule was removed because it has combined with the peptide to form the tetrahedral intermediate.

This reaction intermediate is presumed to be a transient state that will decompose rapidly to form the products that will dissociate from the protease. Once the products are formed, it would be disadvantageous to replace the water and stabilize the products in the binding site. These models for the bound substrate and the tetrahedral intermediate maintained the number and type of atoms and the total charge.

Comparison of minimized structures with the crystal structure of RSV protease

The starting structure consisted of the non-hydrogen atoms from the crystal structure of unliganded RSV protease (Jaskolski et al., 1991) with a model for the missing flap residues in each subunit and for the peptide substrate (Grinde et al., 1992). Previous studies showed that this model was consistent with observed kinetic data, and the structure of the predicted substrate-binding site of RSV protease was very similar to the observed inhibitor binding site of HIV protease (Cameron et al., 1993). The minimized structure for the RSV protease complex with unsubstituted substrate was compared to the starting crystal structure with modeled flaps for the atoms of the protease dimer. The protease-substrate model and starting structure had an RMS deviation of 0.44 Å on C α atoms, 0.67 Å for backbone atoms, 0.78 Å for side-chain atoms, and 0.72 Å for all atoms of the protease. The protease-tetrahedral intermediate model showed similar RMS differences of 0.45 Å on C α atoms, 0.55 Å for backbone atoms, 0.82 Å for side-chain atoms, and 0.69 Å for all atoms, compared to the starting structure.

These values were well within the range of 0.16–0.79 Å for RMS differences observed between main-chain atoms in different crystal forms of the same protein (Wlodawer et al., 1987; Zegers et al., 1994) and close to the mean of 0.40 Å observed for C α atoms (Flores et al., 1993). The differences were spread throughout the structure, except for RMS differences of more than 1 Å for residues 64–66, 70, 71, 64', and 65' in the flaps. This deviation for the flap residues is not surprising because the flaps were not visible in the RSV protease crystal structure. The RMS deviation for side-chain atoms of 0.78–0.82 Å was lower than the range of 1.32–1.68 Å for side-chain atoms in three different crystal forms of bovine pancreatic trypsin inhibitor (Wlodawer et al., 1987). The differences due to minimization were within the range of experimental differences in crystal structures of identical proteins, which indicated the accuracy of the potentials and minimization procedure.

RSV protease was predicted to bind substrate peptides in an extended β -conformation, with each amino acid side chain of the substrate lying in a pocket or subsite formed by the protease residues (Grinde et al., 1992). The substrate peptide that was modeled, Pro-Ala-Val-Ser-Leu-Ala-Met-Thr, consisted of residues P4 to P4', where the peptide bond between Ser at P1 and Leu at P1' is hydrolyzed (Fig. 2). This peptide represented the NC-PR cleavage site in the Gag precursor polyprotein of RSV. The octapeptide was selected as the optimum size for calculations, because tests on HIV protease with decapeptides gave poor results due to larger variation in conformation of the P5 and P5' substrate residues (Weber et al., 1997). The interactions of the modeled peptide with the RSV protease dimer were compared to those of

Fig. 2. Schematic representation of the substrate binding site of RSV protease. The peptide Pro-Ala-Val-Ser-Leu-Ala-Met-Thr that represents the NC-PR cleavage site of RSV is shown. The substrate residues P4–P4' lie in successive protease subsites S4–S4', and the peptide bond between P1 and P1' is hydrolyzed (X). The amino acid substitutions that were tested are indicated at each position from P4–P3'.

the peptide-like inhibitors with HIV protease. Conserved hydrogen bond interactions were observed between HIV protease and the main-chain amides and carbonyl oxygen atoms of the inhibitors in different crystal structures (Gustchina et al., 1994). In the mini-mized RSV protease–substrate models, these conserved β -sheet-like hydrogen bond interactions extended from the carbonyl oxygen of P4 to the amide of P4' (Table 1). The conserved water molecule that was observed in almost all HIV protease–inhibitor crystal structures maintained its four hydrogen bond interactions with the carbonyl oxygens of P2 and P1' of the substrate, and the amides of Ile 67 and 67' in the two flaps. The presence of these conserved interactions again verified the accuracy of the potentials and optimization method.

The model of RSV protease with the tetrahedral intermediate of the substrate showed similar hydrogen bond interactions to those observed in the crystal structures with inhibitors, although the conserved water molecule was no longer present (Table 1). Two exceptions were the carbonyl oxygen of P4, which interacted with both the amide of His 65 and the side chain of Arg 105', and the carbonyl oxygen of P4, which interacted with the side chain of

Arg 10' instead of the amide of Asp 29. One potential interaction between the amide of P3' and the carbonyl oxygen of His 65' did not occur in the model with peptide substrate because the atoms were too far apart. These differences may be due to inaccuracies in the model for the flaps, which included residues His 65 and 65' that were predicted to form several interactions with the substrate peptide, by analogy with the interactions of the equivalent residue Gly 48 in HIV protease (Grinde et al., 1992; Gustchina et al., 1994). In both the model with substrate and the model with the tetrahedral intermediate, the side-chain hydroxyl of Thr at P4' formed a hydrogen bond interaction with the carbonyl oxygen of Pro 62' from the modeled flap. The interactions of the peptides close to the catalytic aspartates differed for the peptide group and the tetrahedral intermediate. The carbonyl oxygen of the scissile peptide bond in the complex of RSV protease with substrate was within hydrogen bond distance of all four δ oxygen atoms of the catalytic Asp 37 and Asp 37' (Table 1). In the tetrahedral intermediate, the carbonyl oxygen was replaced by two hydroxyl groups, which formed hydrogen bond interactions with the δ oxygen atoms of Asp 37.

Substitutions in substrate peptide

Nineteen different substitutions of amino acid residues in the substrate peptide were chosen from those used for kinetic measurements of Cameron et al. (1993). These substitutions were selected to cover the whole range of measured k_{cat}/K_m values (from 25.5 to 0.007 relative to the unsubstituted substrate), and alter residues at positions P4 to P3' (Fig. 2). The kinetic measurements were made on a longer peptide that extended from P5 to P7', but the residues distal to P4-P4' lie outside the protease dimer and were difficult to model because they are expected to have several possible conformations. Therefore, each of the substituted peptides was modeled as an octapeptide in the RSV protease binding site. The different substitutions in the NC-PR peptide and the relative catalytic efficiency, k_{cat}/K_m , for hydrolysis by RSV protease are listed in Table 2. The greatest range of k_{cat}/K_m values was observed for substitutions at P1, whereas less variation was observed in the k_{cat}/K_m values for substitutions at positions further from the scissile bond. Several substitutions of residues P1' to P3' formed peptides that were not hydrolyzed by RSV protease, although hydrolysis was measured using HIV-1 protease (Cameron et al., 1993). The best substrates in this series have large hydrophobic residues, Trp, Tyr, or Phe at P1. In fact, most of the natural cleavage sites of retroviral proteases have medium to large hydrophobic residues at both P1 and P1' (Pettit et al., 1991). Substitutions of Gly at any position gave poor substrates, as did all tested substitutions of P2 and P2' (Cameron et al., 1993). Most of the modeled substrates involved substitutions of hydrophobic residues, but the polar His was modeled at P4 and P3, and Ser at P2 and P2'. Substitutions of other charged residues were not modeled because the neglected effect of solvent was expected to be more pronounced.

Calculated RSV protease–substrate interaction energies

The calculated RSV protease–substrate interaction energies are shown in Table 2 for the 20 modeled peptide substrates and their tetrahedral intermediates. These calculations did not include the energies of solvation, nor entropic terms, but used all-atom potentials with no cutoff limits for the nonbonded terms for increased accuracy. Because the peptides differed by relatively small chem-

Table 1. RSV protease–substrate and HIV protease–inhibitor hydrogen bond interactions^a

Inhibitor or substrate atom	Protease atom	Distance (Å)		
		Crystal ^b	Substrate	Tetrahedral intermediate
P4 O	N His 65	ND ^c	3.4	3.4
P4 O	NH2 Arg 105'		3.4	3.0
P3 N	O δ Asp 41	ND	3.1	3.0
P3 O	N Asp 41	3.0		
P3 O	NH1 Arg 10'		3.3	3.2
P2 N	O His 65	3.0	3.5	3.4
P2 O	O Wat	2.7	3.4	No Wat
P1 N	O Gly 39	3.2		
P1 O(OH)	O δ 1 Asp 37		3.3	3.1
P1 O(OH)	O δ 2 Asp 37		3.0	3.0
P1 O(OH)	O δ 1 Asp 37'		3.2	
P1 O(OH)	O δ 2 Asp 37'		3.0	
P1' O	O Wat	2.7	3.4	No Wat
P2' N	O Gly 39'	3.2		
P2' O	N Asp 41'	3.0	3.0	2.9
P3' N	O His 65'	3.0	3.8 ^d	3.5
P3' O	N His 65'	3.1	3.2	3.0
P4' N	O δ Asp 41'	ND	3.1	3.2
O Wat	N Ile 67	3.1	2.8	No Wat
O Wat	N Ile 67'	3.2	3.3	No Wat
P4' Thr OH	O Pro 62'		2.9	2.9

^aRSV protease–substrate hydrogen bonds are shown for the models with the NC-PR substrate and its tetrahedral intermediate. The P1 O is a carbonyl oxygen in the substrate model and a hydroxyl group (OH) in the tetrahedral intermediate model. The tetrahedral intermediate model does not include the conserved water (Wat) present in the protease–inhibitor crystal structures. The prime indicates a residue in the second subunit of the dimer.

^bCrystal: Mean values calculated for hydrogen bond distances in 15 different HIV protease–inhibitor crystal structures (Gustchina et al., 1994). The equivalent residues in RSV protease are listed.

^cND: Mean value was not calculated (only main chain–main chain interactions were included).

^dDistance too long for hydrogen bond.

Table 2. Minimized RSV protease–substrate complexes

Substitution	k_{cat}/K_m^a	Interaction energy (kcal/mol)		
		Substrate	Tetrahedral intermediate	Difference ^b
None	1.0	−164.1	−276.9	−112.9
P4	His 0.7	−175.4	−278.9	−103.5
	Gly 0.09	−157.6	−262.6	−107.0
P3	Phe 7.5	−177.6	−277.7	−100.1
	His 3.6	−170.4	−297.3	−126.8
	Gly 0.36	−165.0	−273.2	−108.2
P2	Leu 0.06	−168.0	−277.0	−109.1
	Ser 0.01	−160.8	−271.9	−111.1
	Gly 0.04	−154.9	−268.6	−113.7
P1	Trp 25.5	−167.9	−238.6	−70.7
	Tyr 19.1	−170.8	−259.6	−88.8
	Phe 16.7	−172.1	−246.6	−74.6
	Leu 8.5	−162.4	−246.5	−84.1
	Ala 0.5	−157.3	−260.7	−103.4
	Gly 0.1	−153.6	−263.3	−109.7
P1'	Phe 0.4	−166.5	−274.9	−108.4
P2'	Leu 0.1	−137.8	−273.0	−135.3
	Ser 0.01	−170.0	−269.8	−99.9
	Gly 0.007	−161.3	−268.0	−106.7
P3'	Tyr 1.6	−167.2	−272.0	−104.8

^aValues of k_{cat}/K_m are taken from Cameron et al. (1993). The reference substrate is the unsubstituted peptide Pro-Ala-Val-Ser-Leu-Ala-Met-Thr.

^bDifference is the energy difference between the models with the tetrahedral intermediate of the peptide substrate and with the peptide substrate.

ical changes, the differences in solvation energies were expected to be small. Differences in free energy can be related to kinetic measurements by the equation $\Delta G^\ddagger = -RT \ln(k_{cat}/K_m)$ from transition state theory (Fersht, 1985). The velocity of the reaction can be written as $k_{cat}/K_m[E][S]$, and the ratio of the velocities for different substrates at constant $[E][S]$ reflects the difference in the activation energies. The hydrolysis of peptides by RSV protease is a multistep reaction, and the use of $-RT \ln(k_{cat}/K_m)$ can be justified by an extension of the Michaelis-Menten scheme, as described in Fersht (1985, p 102). If the reaction has three steps with rate constants k_s , k_t , and k_p :

where E is enzyme, S is substrate, I is an intermediate, and P is product.

The Michaelis-Menten equation is

$$v = ((k_p k_t)/(1 + k_t)[E_0][S])/(k_s/(1 + k_t) + [S]).$$

By definition, $K_m = (k_p k_t)/(1 + k_t)$, $k_{cat} = k_s/(1 + k_t)$, and k_{cat}/K_m is $(k_p k_t)/k_s$.

Therefore, the differences in free energy are given by

$$\Delta G^\ddagger = -RT \ln(k_{cat}/K_m) = -RT(\ln(k_p) + \ln(k_t) - \ln(k_s)). \quad (1)$$

The $\Delta G^\ddagger$ is linear in the three rate constants, so these terms can be estimated independently. For example, when the final step is slow

and rate limiting, changes in $\Delta G^\ddagger$ or k_{cat}/K_m will reflect changes in k_p . If the reaction has more than one slow step, changes in k_{cat}/K_m will reflect changes in all applicable rate constants.

The differences in free energy $\Delta G^\ddagger = -RT \ln(k_{cat}/K_m)$ were correlated with the calculated protease–substrate interaction energies (Fig. 3). The constant scale factor RT was omitted from the plots. Because the thermodynamic free energy $\Delta G = \Delta H - T\Delta S$, and the calculated interaction energy is an estimate of the internal energy ΔH , correlation will be seen only if the differences in entropy and solvation energies are small.

The statistical significance of the correlation coefficients was assessed with Student's t -test. Because the number of substitutions for each position in the substrate was limited to those for which kinetic data were available, it was necessary to choose a measure that properly treated different numbers of data points. Student's t -test is a valid measure for assessing the significance of correlation, based on the transformation of a correlation coefficient into a Gaussian random variable (Press et al., 1992; Efron & Tibshirani, 1993). The t -test does not assess directly the quality of a linear relationship between data and prediction, but instead gives the probability that a given correlation coefficient could be found by choosing pairs of random numbers. A confidence level of 0.90 means that 9 times out of 10, the random choice would be less correlated than the set being assessed, whereas a confidence level of 0.9995 means that the correlation is highly significant and would arise by chance in only 5 of 10,000 cases. The t -test explicitly includes the dependence of the correlation coefficient on the number of data points. It is inappropriate to test the statistical significance of the parameters of the best-fit line between calculated interaction energy and $\ln(k_{cat}/K_m)$, because these parameters are not expected to be transferable to general binding problems.

The significance of the observed correlation coefficients was evaluated by Student's t -test, as summarized in Table 3. Significant correlations were obtained, despite the fact that both the substrate and the flaps of RSV protease were not visible in any crystal structure. The correlation is best for the set of substitutions of P1 and P1' positions with a value of 79% for the peptide substrate models, 72% for the tetrahedral intermediate models, and 90% for the differences in these two models. These correlation coefficients are comparable in value to those obtained by other methods for inhibitors of HIV protease. Holloway et al. (1995) obtained correlation coefficients (R) of 0.762–0.885 using energy minimization of three training sets of 34 HIV protease complexes with inhibitors differing at P1' and P2'. Head et al. (1996) obtained predictive R values of 0.728 and 0.755 for 13 HIV protease inhibitors. The 14 models with substitutions of P2-P2' positions gave significant correlations of 73% for the tetrahedral intermediate and 71% for the differences, but lower correlation of 36% for the substrate models. The 20 substitutions covering the complete range from P4 to P3' gave the modest correlation coefficient of 46% for the models with peptide substrate. However, this correlation was significant by the Student t -test at the 0.975 level, due to the inclusion of 20 different substrates (Table 3). All 20 models for the tetrahedral intermediates showed a correlation coefficient of 34%, and the Student's t -test suggested that this value may arise by chance (0.90 confidence level). The correlation of 56% for all 20 models of the energy differences between the models with peptide substrate and their tetrahedral intermediate and the observed data was highly significant at the 0.995 level. In fact, the difference between the interaction energies for the bound peptides and their tetrahedral intermediates gave highly significant correlations for all ranges of

Fig. 3. Plot of $-\ln(k_{cat}/K_m)$ against the calculated RSV protease–substrate interaction energy in kcal/mol for substrates with substitutions from P4 to P3'. Values are taken from Table 2. **A:** Models of peptide substrate bound to RSV protease. The line shows the correlation coefficient of 0.793 for the eight peptides with substitutions of P1 and P1' only (solid circles). The correlation coefficient is 0.458 for all 20 substrates (open circles). **B:** Models of the tetrahedral reaction intermediate bound to RSV protease. The line is drawn through the set of eight substitutions of positions P1–P1' (solid circles) and has a correlation coefficient of 0.718. The correlation coefficient is 0.342 for all the substitutions (open circles). **C:** Differences between the calculated values of interaction energy for the peptide–protease complex and the tetrahedral intermediate complex. The line is drawn through the set of eight substitutions of positions P1–P1' (solid circles) and has a correlation coefficient of 0.897. The correlation coefficient is 0.564 for all 20 substrates (open circles).

substitutions with both enzymes. Using the difference in energy changed the reference state for the calculations from the unbound solvated substrate to the bound substrate–enzyme complex. The energy difference will therefore partially cancel effects of solubility differences between free substrates. The difference was a better model for the stabilization of the tetrahedral intermediate by the enzyme. Estimating this stabilization is critical for understanding enzyme-catalyzed reactions.

The substrate models with the worst agreement had Leu or Ser at P2' (Fig. 3A). The substitutions of His at P3, Ser at P2, Ser and Gly at P2' showed the worst agreement for the models of the tetrahedral intermediate (Fig. 3B) and the differences (Fig. 3C). The poor agreement for calculated interaction energies for the models with P2' Ser, Leu, Gly, and P2 Ser may arise from inaccuracies in the kinetic measurements for poor substrates, or from inaccuracies in the model of the RSV protease flaps. In the modeled structure of RSV protease with substrate, the P2 and P2' residues lie close to residues 64–67 of the flap, which were not

visible in the crystal structure (Jaskolski et al., 1991), whereas other positions of the substrate are less close to the flaps (Grinde et al., 1992). P3 His is a relatively large residue in subsite S3 at the protein surface and probably has multiple possible conformations and additional interactions with solvent. Therefore, the conformational entropy will have a larger contribution, resulting in less agreement of the interaction energy (which does not include entropy terms) with the observed catalytic efficiency. In addition, the effects of solvent may be more pronounced, and the calculation does not include a bulk solvent correction, nor an explicit correction for the solvation energies of the uncomplexed protease and peptide.

The reaction mechanism of HIV protease has been studied more extensively than that of RSV protease. However, we assumed that both retroviral proteases follow a similar catalytic reaction. Kinetic analysis has shown that HIV protease-catalyzed hydrolysis can have two different slow steps (Hyland et al., 1991; Polgar et al., 1994). One of the slow steps was proposed to be the binding of

Table 3. Correlation between interaction energies and kinetic measurements^a

Substitutions	Substrate	Tetrahedral intermediate	Difference
P1-P1' (8 peptides)			
RSV PR			
Correlation	0.793	0.718	0.897
Significance	0.990	0.975	0.995
Slope	+0.2639	-0.1115	-0.1148
HIV PR			
Correlation	0.520	0.928	0.964
Significance	0.900	0.9995	0.9995
Slope	+0.2056	-0.1156	-0.1060
P2-P2' (14 peptides)			
RSV PR			
Correlation	0.363	0.733	0.714
Significance	0.90	0.995	0.995
Slope	+0.1178	-0.1757	-0.1220
HIV PR			
Correlation	0.283	0.864	0.888
Significance	0.75	0.9995	0.9995
Slope	+0.0925	-0.1096	-0.1027
P4-P3' (20 peptides)			
RSV PR (20 peptides)			
Correlation	0.458	0.342	0.564
Significance	0.975	0.90	0.995
Slope	+0.1336	-0.0679	-0.0962
HIV PR (21 peptides)			
Correlation	0.084	0.640	0.665
Significance	0.60	0.995	0.9995
Slope	+0.0163	-0.0846	-0.0815

^aThe correlation coefficient $R = \frac{\sum(x - \langle x \rangle)(y - \langle y \rangle)}{(\sum(x - \langle x \rangle)^2 \times \sum(y - \langle y \rangle)^2)^{1/2}}$ where x and y are the variables, and $\langle x \rangle$ and $\langle y \rangle$ are the mean values of x and y . Significance is estimated from the Student's t -test. Slopes of the lines of best fit to data in Table 2 are shown. The correlation is calculated between $-\ln(k_{cat}/K_m)$ and the calculated RSV protease-substrate interaction energy, using the values in Table 2 for bound peptide substrate, tetrahedral intermediate of substrate, and the difference in energy between the peptide substrate and its tetrahedral intermediate. HIV protease values are from Weber and Harrison (1996). Some of the peptides differ in the calculations for RSV and HIV proteases.

substrate to the enzyme, and the other involved proton transfer, which must occur on formation or decomposition of the tetrahedral intermediate. Because there were two slow steps, there cannot be a unique transition state (highest energy point on the reaction coordinate) for this enzyme. In other words, the tetrahedral intermediate is not always the transition state. The observed $\Delta G^\ddagger$ values correlated well with the calculated interaction energies for both the peptide substrates and their tetrahedral intermediates when the substitutions are in positions P1 and P1' (Table 3). For all 20 peptides, the differences between the calculated interaction energies for tetrahedral intermediate and peptide substrate provided a better estimate than either energy alone, although the protease-substrate energies were significant at the 0.975 level (Table 3). These significant correlations suggested that the catalytic reaction of RSV protease had slow steps for both the binding of substrate and the stabilization of the tetrahedral intermediate (Fig. 4). In contrast, when similar calculations were run on HIV-1 protease, little or no correlation was seen between the calculated substrate

binding energy and the observed kinetic data, whereas the models for the tetrahedral intermediate showed significant correlations. This result suggested that the dissociation of the tetrahedral intermediate was the major slow step in the reaction for HIV protease (Weber & Harrison, 1996). The lower correlations observed for RSV protease may be due to inaccuracies in the model for the protease structure, because the crystal structure did not show the flap residues or an inhibitor. The more significant contribution from the first step of binding substrate to RSV protease can be explained by the higher energy needed to stabilize the longer flaps, which included a three-residue insertion compared to those of HIV protease, and are expected to be more flexible. Alternatively, the lower correlation for models of RSV protease may arise from a greater variation in the entropic contribution of different substrates compared to the entropic contribution with HIV protease.

Fig. 4. Reaction coordinates consistent with experimental data for HIV protease catalysis from enzyme (E) and substrate (S) via intermediate (I) to enzyme and products (P) are shown. Two possible slow steps (Hyland et al., 1991; Polgar et al., 1994) are indicated by the height of the energy barriers and labeled ET. **A:** Binding of the substrate peptide is the slow step and the calculated energy for ES (dashed lines) would correlate with the rate. **B:** Dissociation of the tetrahedral intermediate is the slow step and the energy of EI (dashed lines) would anti-correlate with the rate.

Discussion of the entropic contribution

The interaction energies calculated from molecular mechanics are usually much larger in magnitude than the observed free energy differences. It can be shown that this is the result of using a single point estimate for a thermodynamic average. Molecular mechanics and dynamics calculations are valuable for the analysis of experimentally determined structures. However, these calculations are a simplistic classical approximation to an unsolvable quantum problem, and even a long (>1,000 ps) molecular dynamics simulation is an insufficient statistical sample with respect to the behavior of 10^{12} or more molecules over seconds to hours of time. The solution is to include as much experimental data as possible, and to estimate the effects of small perturbations. Molecular mechanics calculations provide a point estimate of the internal energy for a specific molecular conformation. The free energy is determined by the partition function, which is an ensemble average of these point estimates.

The effect of using one point in the ensemble can be treated analytically. The interpretation of molecular mechanics internal energies is best understood in terms of Feynmann's minimum theorem (Feynmann, 1972). This theorem is derivable from information theory and is known in other contexts as a Chernoff bound (Harrison, 1986; Evertsz & Mandelbrot, 1992). It relates differences in the total energy of a system to differences in the free energy. In statistical mechanics, the Helmholtz free energy, F , is defined by the log of the partition function:

$$F = -RT \ln Z.$$

The partition function, Z , is the sum of the Boltzmann weights for all of the relevant states of the system:

$$Z = \Lambda^{3N/2} \sum e^{-H/RT},$$

where R is the ideal gas constant, T the temperature, N is the number of atoms, Λ is a volume normalizing factor, and H is the Hamiltonian or total energy (molar units are used in these equations). The sum is over the "states" of the system, which, for this problem, is the range of conformations or configurations of atoms.

The first term in a power series expansion of F in terms of a perturbation in H for one given configuration results in a linear expansion. The value of ΔH is estimated from the molecular mechanics calculation for this configuration.

$$F = F_0 + dF/dH \Delta H$$

$$\Delta F = F - F_0$$

$$\Delta F = -RT/Z(dZ/dH)\Delta H$$

$$\Delta F = -RT/Z(-\Lambda^{3N}e^{-H/RT}/RT)\Delta H$$

$$\Delta F = (\Lambda^{3N}e^{-H/RT}/Z)\Delta H.$$

The molecular mechanics approximation of using a conformation near an energy minimum will be most accurate when one conformation dominates the partition function because the ratio $\Lambda^{3N}e^{-H/RT}/Z$ will be near unity. The difference in the Hamiltonian (ΔH) will always be larger than the difference in the free energy ($\Delta \ln Z$), because the linear expansion for $\ln(Z)$ is $Z - 1$ and $Z - 1$ is always equal to or greater than $\ln(Z)$. Consequently, molecular mechanics calculations will have an apparent ideal gas constant

that is different from R . Therefore, the slope for the line of best correlation between $-\ln(k_{cat}/K_m)$ and the calculated interaction energy (Fig. 3; Table 4) is not equal to RT . Because the entropy is $R \ln Z$, the slope, $(\Lambda^{3N}e^{-H/RT}/Z)$, gives an estimate of the average entropy of binding. This effect is similar to the arbitrary scales in linear free energy calculations such as the Hammett equation (March, 1992, pp 278–286). The measured differences in free energy are lower than the differences in calculated interaction energy.

The entropic contribution to peptide hydrolysis by RSV or HIV proteases can be estimated from the slope of the line of regression for the calculated interaction energies and $-\ln(k_{cat}/K_m)$ from kinetic measurements. These slopes are listed in Table 3 for correlations observed for RSV protease (Fig. 3) and HIV protease (Weber & Harrison, 1996). It is only valid to compare the slopes where the calculations showed significant correlation for both enzymes. For the differences in energy between the tetrahedral intermediate and substrate, the slopes for models of RSV and HIV proteases were very similar, with values of -0.08 to -0.12 . For the tetrahedral intermediates, the slopes had a slightly greater range of values, from -0.07 to -0.18 . The similar values for slopes suggested that the entropic contribution to binding was similar for the hydrolysis of the NC-PR series of peptides by both enzymes. Because the slopes were similar, the lower correlations in the RSV protease were probably due to inaccuracies in the model for the flaps and the substrate, especially close to P4, P3, and P3'.

Molecular mechanics calculations of the interaction energies gave significant correlation with free energy differences derived from kinetic data for 20 RSV protease–substrate complexes with substitutions of mostly hydrophobic residues in the substrates. These calculations were performed with a transferable molecular mechanics potential that did not include empirical corrections, empirical ligand binding terms, or solvent corrections. Interestingly, comparison with similar calculations on HIV protease complexes suggested that the slow step for hydrolysis of substrates may differ for the two enzymes. However, the entropic contribution, as estimated from the slopes of the plots of calculated and experimental free energy differences, was similar for hydrolysis by both RSV and HIV proteases. These calculations showed promise for predicting the specific changes in peptides that will form better or worse substrates of retroviral proteases. The molecular mechanics calculation can also be applied to design improved inhibitors of HIV protease.

Materials and methods

Initial model

The starting structure consisted of the protease dimer from the crystal structure of RSV protease (Jaskolski et al., 1991). The flap residues 61–69 from one subunit and 59–69 from the second subunit were not visible in the electron density map, so these residues were modeled as described (Grinde et al., 1992). The starting model for the central eight residues, Pro-Ala-Val-Ser-Leu-Ala-Met-Thr, of the peptide substrate was obtained by mutating the side chains in a peptidic inhibitor and forming a peptide group (CONH) instead of the nonhydrolyzable bond of the inhibitor, as described for HIV protease (Harrison & Weber, 1994). The structures were examined on a Silicon Graphics Indigo 2 using the program CHAIN (Sack, 1988). The protease–substrate complex was modeled with a proton (H^+) positioned midway between the closest side-chain oxygens of the two adjacent catalytic aspartates, Asp 37 and

Asp 37', and the carbonyl oxygen of the scissile peptide bond. The catalytic aspartic acids were negatively charged. The position of this proton has been shown to be energetically stable by molecular dynamics simulations of HIV protease with substrate (Harrison & Weber, 1994). A tetrahedral intermediate of the substrate was also modeled in which the scissile peptide CO-NH was replaced by $[C(OH)_2-NH_2]^+$ and the proton and the water that interacts with the substrate or inhibitor and the flaps of the protease were removed. This model for the tetrahedral intermediate maintains the number and type of the atoms and the total charge. The tetrahedral intermediate was chosen to represent the reaction intermediate in the AAC2 reaction mechanism for hydrolysis of amides (March, 1992, p 385). This reaction mechanism is the predominant form seen in acid-catalyzed hydrolysis of amides in model systems. In this reaction mechanism, the carbonyl oxygen is protonated and the activated carbo-cation is attacked by an unactivated water molecule, resulting in a charged tetrahedral intermediate. The attack by water is normally the slow step in the reaction. Both the bound substrate and the tetrahedral intermediate can readily be modeled from the structures of peptidic inhibitor complexes with HIV protease.

Energy minimization calculation

The minimization and molecular dynamics were run using the program AMMP (Harrison, 1993), which incorporates a new algorithm to increase the speed and allow calculation of nonbonded and electrostatic terms without use of a cutoff radius. The molecular mechanics calculations used the UFF potential set (Rappe et al., 1992) with parameters modified as described in Weber and Harrison (1996). In addition, the force constants for the planarity terms of the carbonyl and aromatic carbon atoms were increased from 6 kcal/mol-Å to 150 and 100 kcal/mol-Å, respectively. This change significantly improved the agreement between the calculated normal mode frequencies and observed infrared absorptions for formaldehyde and benzene. Charges for the nonstandard groups in the tetrahedral intermediate were generated as described for HIV-1 protease (Weber & Harrison, 1996). The charges were generated for each tetrahedral intermediate in the complex with RSV protease before the minimization. The positions for the non-hydrogen atoms of the protein were fixed initially in order to calculate and minimize the hydrogen atom positions. No screening dielectric term or bulk solvent correction was included. A constant dielectric of 1 was used. No cutoff was applied for nonbonded and electrostatic terms. These terms were calculated with an algorithm that amortizes or spreads the cost of calculation over many simpler calculations, which results in lower average cost, as described in Harrison and Weber (1994). This algorithm, when used in conjunction with the fast multipole method as implemented in AMMP, brings the cost of the calculation without cutoffs below the cost of using a cutoff of 6–8 Å with standard approaches.

The RSV protease structure with non-hydrogen atoms from the crystal structure, modeled flaps, and minimized hydrogen atoms was combined with the RSV NC-PR peptide substrate that had been minimized in a complex with HIV-1 protease (Weber & Harrison, 1996). Each of the side-chain torsion angles for residues in the peptide substrate was rotated through 360° in steps of 15° to search for alternate conformations of these residues. This torsion search finds the angle(s) that have a minimum in the nonbonded and electrostatic energy terms. The modeled complex was optimized by a longer minimization using 100 steps of conjugate gra-

dients followed by 8 cycles of alternating 30 steps of conjugate gradients and short runs of molecular dynamics (20 steps of 1fs at 300K). The molecular dynamics serves to introduce a small random "error" into the coordinates and improves the overall convergence. The same protease structure was combined with each of the different peptides with single amino acid substitutions. The new atoms of each substituted amino acid side chain of the substrate were built with AMMP in the model of the protease-peptide complex. The side-chain atoms were removed down to the C β atom, or the C α for substitution of Gly, and the new atomic positions created by a variation on distance geometry (Harrison et al., 1995). The new atoms were minimized with respect to bond, angle, torsion, and hybrid potentials. The torsion angles of the new side chain were searched for the minimum energy conformation. Then each model of RSV protease with a different substrate was optimized by the longer minimization as described above. The tetrahedral intermediate was made for each substrate by deleting the carbonyl oxygen and building the new atoms of the tetrahedral intermediate as described for the new side-chain atoms. Then, the same longer minimization procedure was used for the protease-tetrahedral intermediate complex. Finally, the protease-substrate nonbonded interaction energy was calculated for each of the optimized structures of RSV protease with different substrates and their tetrahedral intermediates.

Acknowledgments

We thank Jozsef Tozsef and Jonathan Leis for valuable discussions. This work was supported in part by NIH grant CA58166.

References

- Cameron CE, Grinde B, Jacques P, Jentoft J, Leis J, Wlodawer A, Weber IT. 1993. Comparison of the substrate binding pockets of the Rous sarcoma virus and human immunodeficiency virus type 1 proteases. *J Biol Chem* 268:11711–11720.
- Cameron C, Grinde B, Jentoft J, Leis J, Weber I, Copeland T, Wlodawer A. 1992. Mechanism of inhibition of the retroviral protease by a RSV peptide substrate representing the cleavage site between the gag p2 and p10 proteins. *J Biol Chem* 267:23735–23741.
- Efron B, Tibshirani RJ. 1993. *An introduction to the bootstrap. Monographs on statistics and applied probability* 57. New York: Chapman & Hall. pp 53–56; 156–159.
- Evertsz CJG, Mandelbrot BB. 1992. *Multifractal measures*. In: Peitgen HO, Jurgens H, Saupe D, eds. *Chaos and fractals*. New York: Springer-Verlag. pp 948–949.
- Fersht A. 1985. *Enzyme structure and mechanism*. New York: W.H. Freeman and Company.
- Feynmann RP. 1972. *Statistical mechanics, a set of lectures*. Redwood City, California: Addison-Wesley.
- Flores TP, Orengo CA, Moss DS, Thornton JM. 1993. Comparison of conformational characteristics in structurally similar protein pairs. *Protein Sci* 2:1811–1826.
- Fruton JS. 1987. The mechanism of the catalytic action of pepsin and related acid proteinases. In: Neuberger A, Brocklehurst K, eds. *Hydrolytic enzymes*. Amsterdam, The Netherlands: Elsevier Science Publishers. pp 1–37.
- Grinde B, Cameron CE, Leis J, Weber I, Wlodawer A, Burstein H, Skalka AM. 1992. Analysis of substrate interactions of the Rous sarcoma virus and human immunodeficiency virus-1 proteases using a set of systematically altered peptide substrates. *J Biol Chem* 267:9491–9498.
- Gustchina A, Sansom C, Prevost M, Richelle J, Wodak SY, Wlodawer A, Weber IT. 1994. Energy calculations and analysis of HIV-1 protease-inhibitor crystal structures. *Protein Eng* 7:309–317.
- Harrison RW. 1986. Sayre's equation is a Chernov bound to maximum entropy. *Acta Crystallogr A* 43:428–430.
- Harrison RW. 1993. Stiffness and energy conservation in the molecular dynamics—An improved integrator. *J Comp Chem* 14:1112–1122.
- Harrison RW, Chatterjee D, Weber IT. 1995. Analysis of six protein structures predicted by comparative modeling techniques. *Proteins Struct Funct Genet* 23:463–471.

- Harrison RW, Weber IT. 1994. Molecular dynamics simulation of HIV-1 protease with peptide substrate. *Protein Eng* 7:1353–1363.
- Head RD, Smythe ML, Oprea TI, Waller CL, Green SM, Marshall GR. 1996. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J Am Chem Soc* 118:3959–3969.
- Holloway MK, Wai JM, Halgren TA, Fitzgerald PM, Vacca JP, Dorsey BD, Levin RB, Thompson WJ, Chen LJ, deSolms SJ, Gaffin N, Ghosh AK, Giuliani EA, Graham SL, Guare JP, Hungate RW, Lyle TA, Sanders WM, Tucker TJ, Wiggins M, Wiscount CM, Woltersdorf OW, Young SD, Darke PL, Zugay JA. 1995. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J Med Chem* 38:305–317.
- Hyland LJ, Tomaszek TA, Meek TD. 1991. Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry* 30:8454–8463.
- Jaskólski M, Miller M, Mohana Rao JK, Leis J, Wlodawer A. 1991. Structure of the aspartic protease from Rous sarcoma retrovirus refined at 2 Å resolution. *Biochemistry* 29:5889–5898.
- Jeffrey GA, Saenger W. 1991. *Hydrogen bonding in biological structures*. Berlin: Springer-Verlag.
- Kourinov I, Harrison RW. 1994. Prediction of novel serine proteinase inhibitors. *Nature Struct Biol* 1:735–743.
- March J. 1992. *Advanced organic chemistry*, 4th ed. New York: J. Wiley & Sons.
- Miller M, Jaskólski M, Mohana Rao JK, Wlodawer A. 1989. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* 337:576–579.
- Oroszlan S, Luftig RB. 1990. Retroviral proteinases. *Curr Topics Microbiol Immunol* 157:153–185.
- Pettit SC, Simsic J, Loeb DD, Everitt L, Hutchinson CA III, Swanstrom R. 1991. Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. *J Biol Chem* 266:14539–14547.
- Polgar L, Szeltner Z, Boros I. 1994. Substrate-dependent mechanisms in the catalysis of human immunodeficiency virus protease. *Biochemistry* 33:9351–9357.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BR. 1992. *Numerical recipes in C*, 2nd ed. Cambridge: Cambridge University Press. pp 636–639.
- Rappe AK, Casewit CJ, Colwell KS, Goddard WA III, Skiff WM. 1992. UFF: A full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc* 114:10024–10035.
- Sack JS. 1988. CHAIN—A crystallographic modeling program. *J Mol Graph* 6:224–225.
- Smith R, Brereton IM, Chai RY, Kent SBH. 1996. Ionization states of the catalytic residues in HIV-1 protease. *Nature Struct Biol* 3:946–950.
- Tozser J, Bagossi P, Weber IT, Oroszlan S. 1996. Comparative studies on the substrate specificities of avian myeloblastosis virus proteinase and lentivirus proteinases. *J Biol Chem* 271:6781–6788.
- Weber IT. 1990. Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and Rous sarcoma virus proteases. *J Biol Chem* 265:10492–10496.
- Weber IT, Cavanaugh DS, Harrison RW. 1997. Models of HIV-1 protease with peptides representing its natural substrates. *J Mol Struct (Theochem)*. Forthcoming.
- Weber IT, Harrison RW. 1996. Molecular mechanics calculations on HIV-1 protease with peptide substrates correlate with experimental data. *Protein Eng* 9:679–690.
- Wlodawer A, Erickson JW. 1993. Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem* 62:543–585.
- Wlodawer A, Miller M, Jaskólski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SBH. 1989. Conserved folding in retroviral proteases—Crystal structure of a synthetic HIV-1 protease. *Science* 245:616–621.
- Wlodawer A, Nachman J, Gilliland GL, Gallagher W, Woodward C. 1987. Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J Mol Biol* 198:469–480.
- Xu LZ, Weber IT, Harrison RW, Gidh-Jain M, Pilgis SJ. 1995. Sugar specificity of human β -cell glucokinase—Correlation of molecular models with kinetic measurements. *Biochemistry* 34:6083–6092.
- Zegers I, Maes D, Dao-Thi MH, Poortmans F, Palmer R, Wyns L. 1994. The structures of RNase A complexed with 3'CMP and dCpA—Active site conformation and conserved water molecules. *Protein Sci* 3:2322–2339.